429—435　　3502(15)

动 物 学 研 究 1997, 18 (4): 429—435　　CN 53-1040/Q ISSN 0254-5853
Zoological Research

从氨基酸残基的静态可及性探讨组蛋白的分子进化*

黄京飞　刘次全
(中国科学院昆明动物研究所细胞与分子进化开放实验室　昆明　650223)

A **摘要**　本文以蛋白质氨基酸残基的静态可及性为指标，分别探讨了 H_{2A}、H_{2B}、H_3、H_4 这4种组蛋白分子的进化关系，并获得了 H_{2A}、H_{2B} 和 H_3 各自相应的分子系统树。该研究从分子进化的角度表明，在构成核小体的这4种组蛋白分子中，组蛋白 H_4 的保守性最大。作者推测，与其他几种组蛋白相比，在维持染色体的结构、功能以及在染色体的起源上，组蛋白 H_4 可能具有最为重要的意义和作用。

关键词　组蛋白，静态可及性，分子进化，三维结构　氨基酸残基，

分子进化的研究表明，在分子水平上，蛋白质的三维结构可为研究其进化关系提供非常有用的信息，在进化差异上有作用的绝大多数约束都体现在三维结构水平上，而且在进化上，生物大分子的三维结构具有比其序列更大的保守性(Bajaj 等，1984；Johnson 等 1990)。

由于球蛋白的天然结构仅存于水溶液环境中，因而对其与水相互作用的分析是蛋白质结构理论的核心(Kauzmann，1959)。为了研究蛋白质与水的相互作用，Lee 等(1971)引入了"可及表面"(或称"静态可及性")的概念，对与水接触的蛋白质表面原子面积的大小进行定量描述。蛋白质的表面可及性不仅可以反映蛋白质与其他分子的接触程度、蛋白质分子的面积、体积等，而且还可为确定蛋白质与其他分子发生反应的位置提供有意义的信息。因此，一些人就蛋白质表面可及性的计算、可及性与蛋白质折叠的关系以及通过研究这种关系来大概地预测蛋白质的三维结构等方面进行了一系列有意义的研究。

目前普遍认为，组蛋白分子在进化上具有极大的保守性，而且在真核生物的染色体中，由4种组蛋白分子八聚体所构成的核心颗粒是染色体结构的的基本单元——核小体的重要组成部分。因此，有关组蛋白结构与分子进化的研究，特别是对组蛋白分子三维结构与进化关系的研究，将会对染色体结构、功能与进化关系以及分子遗传学中若干重要问题的研究起到积极的推动作用。然而，有关组蛋白三维结构与其进化关系问题的研究，至今

---

* 云南省应用基础研究科学基金资助项目
本文1996年3月7日收到，同年6月6日修回

仍未见报道。这主要是由于技术方面的原因，目前人们仍无法获得组蛋白分子的晶体结构数据。可及性的变化可以为从一级结构推测三级结构提供一个量的依据(Lee 等，1971；Richards，1977)，因此，可及性中包含着蛋白质三维结构的某些信息。此外，表面可及性已被广泛用于研究单个氨基酸残基在蛋白质三维结构中的溶剂可及性。一般来讲，溶剂不可及的残基的突变率比那些位于分子表面的溶剂可及的残基有较低的突变率(Chothia 等，1986；Hubbard 等，1987；Lim 等，1989)；不同可及性氨基酸残基的可变性和保守性大不相同，溶剂不可及的残基通常较为保守(Overington 等，1990，1992；Topham 等，1993)，因此，可及与不可及残基替换模式的不同反映了这两类残基保守性的差异(Wako 等，1994)。基于上述观点，本文利用组蛋白分子在伸展和天然两种状态下氨基酸分子的静态可及性，就 H_{2A}、H_{2B}、H_3 和 H_4 4 种类型组蛋白分子的系统演化关系进行了探讨。

1 方法

首先，我们利用蛋白质氨基酸残基在伸展和天然折叠状态中的可及性值(Chothia，1976)，通过下式计算出各种组蛋白分子相互间在静态可及性上的相似度($R_{i,j}$)，然后，通过对所获得的 $R_{i,j}$ 值进行模糊聚类分析(黄京飞等，1991)，便可获得相应的系统树。

$$R_{i,j}=\frac{\left|\sum_{k=1}^{N}A_{k,i}A_{k,j}\right|}{\left(\sum_{k=1}^{N}A_{k,i}^{2}\sum_{k=1}^{N}A_{k,j}^{2}\right)^{1/2}} \tag{1}$$

式中，$A_{k,i}$ 和 $A_{k,j}$ 分别表示组蛋白序列 i 和 j 中的第 k 个氨基酸在伸展或折叠状态下的静态可及性值，N 为序列经对齐处理后的总长度。本研究所涉及的血红蛋白和肌红蛋白序列，分别采用 Creighton(1983) 及 Johnson 等 (1990) 的对齐处理后的形式，而所有组蛋白序列则采用 Wells(1986) 所收集和对齐处理后的排列形式。

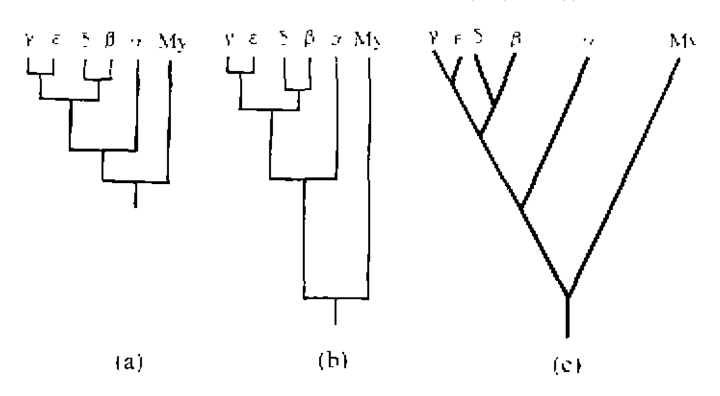


图 1 血红蛋白 α、β、γ、δ、ε 链与肌红蛋白间的进化关系

Fig.1 The evolutionary scheme between the α、β、γ、δ and ε chain of hemoglobin and human myoglobin

a. 肽链在伸展状态下，以分子静态可及性相似度为指标构建的系统树；b. 肽链在折叠状态下，以分子静态可及性相似度为指标构建的系统树；c. 根据比较基因和蛋白序列的相似性而获得的系统树(Efstratiadis 等，1980)。图中，α、β、γ、δ 和 ε 分别为血红蛋白的 α、β、γ、δ 和 ε 链；My 为人的肌红蛋白。

a. The phylogenetic tree is constructed by similarities between amino acid static accessibilities in extending state

of peptide chain; b The tree is constructed by similarities between amino acid static accessibilites in folding state of peptide chain; c The tree inferred from the relative similarities of their gene and protein sequences (Efstratiadis *et al*, 1980) In the figures, α, β, γ, δ and ε represent hemoglobin α, β, γ, δ and ε chain respectively, My is human myoglobin.

2 结果

为了对方法的可靠性进行检验，我们对血红蛋白和肌红蛋白进行了研究，得到了相应的分子系统树(图 1: a, b 和图 2: a, b)，并将其与通过一维和三维结构比较所获得的系

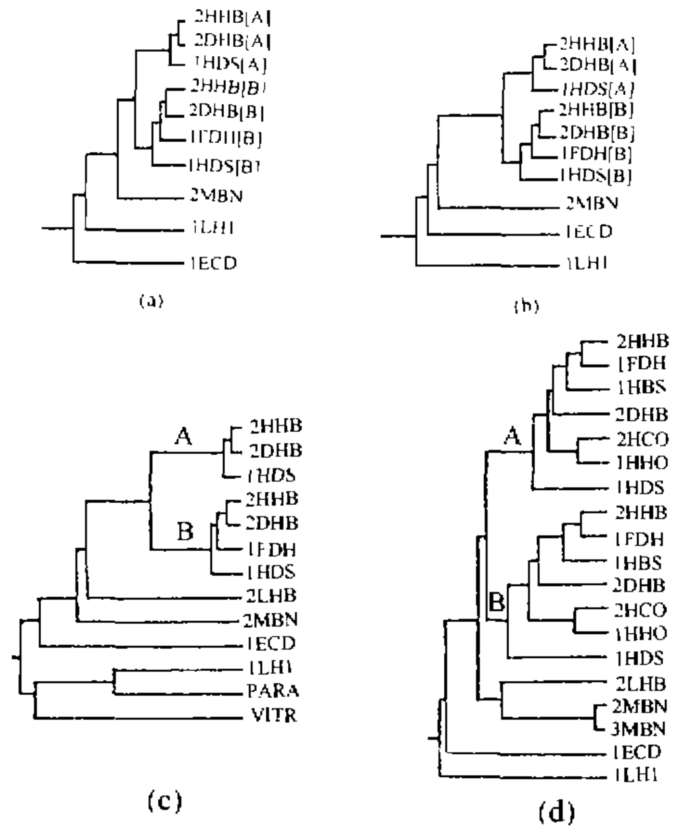


图 2　血红蛋白与肌红蛋白的分子系统树

Fig. 2　The phylogenetic trees of hemoglobins and myoglobin

a 肽链在伸展状态下，以分子静态可及性相似度为指标构建的系统树；b 肽链在折叠状态下，以分子静态可及性相似度为指标构建的系统树；c. 通过比较氨基酸序列而获得的系统树(Johnson 等, 1990)；d. 通过直接比较蛋白主链的拓扑结构而获得的系统树(Johnson 等, 1990)。

a The phylogenetic tree is constructed by similarities between amino acid static accessibilities in extending state of peptide chains; b. The phylogenetic tree is constructed by similarities between amino acid static accessibilities in folding state of peptide chains, c. The tree for the globins based sequence distances (Johnson *et al.*, 1990); d The tree based on the 3-D structural distances (Johnson *et al*, 1990). In the figures, 2HHB: Human deoxy α- and β-chains-hemoglobin, 2HHO: Human carbonmonoxy α- and β-chains-hemoglobin; 1HHO: Human oxy γ-and β-chains-hemoglobin, 1HBS Human Sickle-cell hemoglobin α- and β-chains, 1FDH Human deoxy α- and γ-fetal chains-hemoglobin, 2DHB: Horse deoxy α- and β- chains-hemoglobin; 1HDS: Deer sickle cell hemoglobin γ- and β-chains; 1LHB: Sea lamprey hemoglobin V (cyano.' met). 2MBN: Sperm whale met myoglobin. 3MBN. Sperm whale deoxy myoglobin; 1ECD *Chironomus thummi thummi* erythrocruorin, 1LH1: *Lupinus luteus* leg hemoglobin

统树(图1c和图2：c，d)分别进行了对比。结果显示，利用可及性为指标建立的系统树与利用序列和三维结构比较所得系统树，基本上是一致的。

在分别计算了4种类型组蛋白分子在伸展和折叠状态下相互间在可及性方面的相似度之后，我们得到了H_{2A}、H_{2B}和H_3各自的分子系统树(图3：a-f)。结果表明，以折叠状态下的$R_{i,j}$值为指标构建的系统树(图3：d，e，f)，优于用伸展状态下的这一值建立的系统树(图3：a，b，c)。此外，我们通过计算分子间的疏水相似性(黄京飞等，1996)以及运用模糊聚类分析法，分别对H_{2A}、H_{2B}、H_3进行序列分析后建立了相应的系统树。而对于H_4，我们使用的方法不能获得理想的系统树。通过对H_4的氨基酸序列进行分析后，可以看出，这主要是由于H_4的序列在进化上极度保守所致。在我们选用的从酵母、海胆直至哺乳类动物的8个物种中，竟有6个序列完全相同，而这6个都属于脊椎动物；即使是海胆与脊椎动物的序列相比，在全长102个氨基酸残基的序列中，也仅差1个氨基酸。

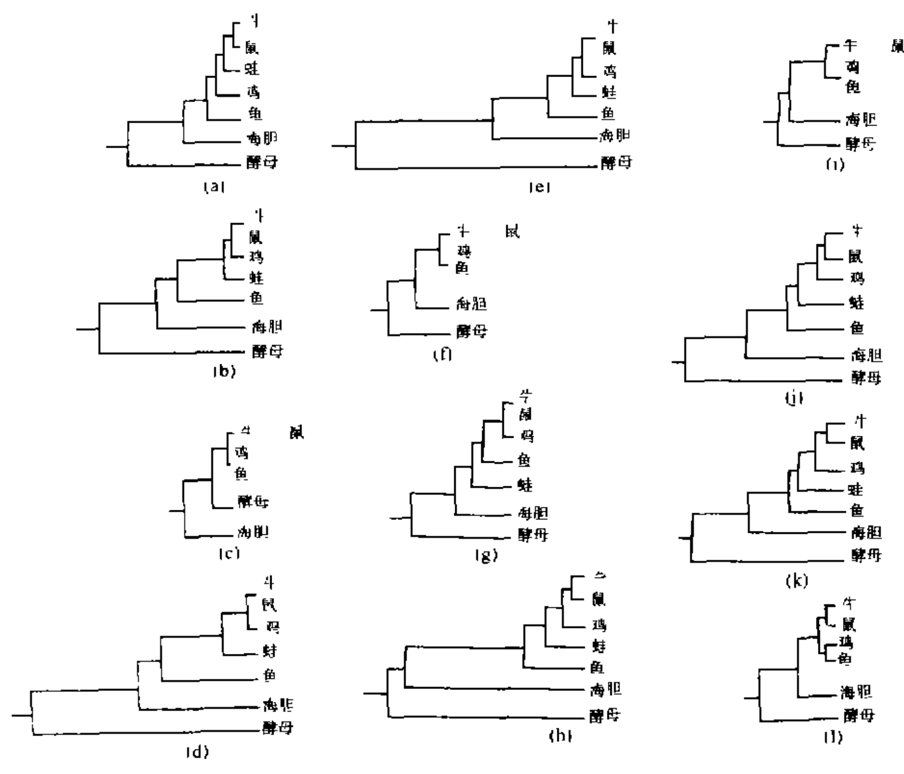


图3 组蛋白的分子系统树

Fig. 3 The phylogenetic trees of histones

a、b、c分别为H_{2A}、H_{2B}、H_3的肽链在伸展状态下，以分子间静态可及性上的相似度为指标构建的系统树；d、e、f分别为H_{2A}、H_{2B}、H_3的肽链在折叠状态下，以分子间静态可及性上的相似度为指标构建的系统树；g、h、i分别为利用H_{2A}、H_{2B}、H_3的疏水相似性为指标建立的系统树；j、k、l为运用模糊聚类分析法分别对H_{2A}、H_{2B}、H_3进行序列分析后建立的系统树。

a, b and c are the phylogenetic trees constructed by similarities between amino acid static accessibilities of H_{2A}, H_{2B} and H_3 in extending state of peptide chains; d, e and f are the trees of H_{2A}, H_{2B} and H_3 in folding state of peptide chains; g, h, i are the trees of H_{2A}, H_{2B} and H_3 constructed by molecular hydrophobic similarities; j, k, l are the trees of H_{2A}, H_{2B} and H_3 constructed by the fuzzy cluster method comparing histone sequences.

3 讨论

生物大分子的三维结构不仅在进化上具有比其一级结构更大的保守性，而且三维结构和生物大分子的进化同具非线性特征，因而，从非线性理论的认识论出发，就蛋白质的非线性三维结构与分子进化的关系进行研究，对于探索生物大分子进化的复杂性与非线性等理论生物学中的重大问题，无疑具有十分重要的意义。

由于蛋白质的静态可及性在一定程度上反映了蛋白质三维结构的某些信息，因而，通过比较其分子间静态可及性的相似度，也可间接地反应它们三维结构间的异同，尤其对那些未能获得三维结构晶体数据的蛋白质分子(如组蛋白)，这种方法就更具实际应用价值。从研究结果来看，根据比较组蛋白在折叠状态下氨基酸的静态可及性所获得的系统树(图3：d、e、f)较为合理，它们与用其他一些方法所获得的结果是一致的(图3：g—l)；而通过计算组蛋白在伸展状态下氨基酸残基的静态可及性所得到的系统树(图3：a、b、c)与前者相比，其合理性则稍差，如在图2a中的1ECD和1LH1以及H_{2A}中的蛙与鸡(图3：a、d)以及H_3中的酵母与海胆分别在两种系统树中的位置都有差异(图3：c、f)。这是由于以伸展状态下氨基酸残基的可及性所建立起来的系统树，实质上仍旧是对一级结构间的异同进行比较的结果。而当考虑折叠状态下的可及性时，在我们的方法中虽然仍是以序列为基础，但这时已包含了三维结构的因素，尽管它只是一种粗略和间接的。因此，我们认为，上述两种系统树的异同，从另一个侧面表明了，生物大分子的线性一级结构与非线性三维结构在进化过程中的联系与差别。

在结构方面，H_{2A}、H_{2B}、H_3和H_4这4种类型的组蛋白分子都有较大的保守性，但比较而言，在它们当中以H_4的保守性最大，以致在脊椎动物中，它的结构几乎没有任何差别；H_3的保守性次之，因为在脊椎动物中，哺乳类与其他脊椎动物H_3的结构仍有所不同；H_{2A}和H_{2B}二者的结构保守性基本一致。研究结果表明，对于这4种组蛋白在每亿年里每100个氨基酸中的替换率分别为：$H_{2A}=1.7$，$H_{2B}=1.7$，$H_3=0.30$，$H_4=0.25$ (Efstratiadis 等，1980)。我们的结果与之完全吻合。作者认为，从分子进化研究的角度来讲，如对各类真核生物组蛋白间的系统演化关系进行研究，则选用H_{2A}和H_{2B}似乎较为合适；但如只探讨哺乳类或脊椎动物与其他真核生物间的关系，也可选用H_3或H_4。由于组蛋白结构具有高度的保守性，它仅适用于对进化距离较远的生物类群进行分子进化的研究，而当用组蛋白作为材料来对进化距离较近的类群进行分子进化的研究时，则可能会出现较大的误差。

需要特别指出的是，组蛋白H_4由于结构的极端保守性，用现行的所有研究分子进化的方法都很难得到其理想的系统树。但H_4在维持染色体的基本结构和功能方面具有相当重要的作用，以致在其序列中，每个氨基酸残基的改变，都有可能对其结构和功能产生较大的影响，因而，H_4的结构在进化过程中几乎没有变化。由此，我们可以推测，组蛋白H_4可能是染色体在进化过程中最早具有的碱性蛋白质。实验表明，在原核与真核生物的过渡类型——甲藻中，已发现其染色体中具有与小牛胸腺组蛋白H_4相似的碱性蛋白(李靖炎，1979)。因此，作者认为，与其他种类的组蛋白相比，H_4在维持染色体的结构和功能上也许是最为重要的，在染色体的起源上也可能占据着重要的位置。

致谢 本文承蒙李靖炎先生提出宝贵意见，并与作者进行了有益的讨论，谨此致谢。

参 考 文 献

李靖炎，1979 细胞在生命进化历史中的发生——真核细胞的起源. 北京：科学出版社. 40—58.

黄京飞、刘次全、王 莹，1991 分子进化研究：I. 细胞色素C分子进化的模糊聚类分析法研究. 动物学报，**37**(4)：443—451.

黄京飞，刘次全，1996. 细胞色素分子疏水性与进化的关系. 动物学研究，**17**(2)：179—185.

Bajaj M，Blundell T. 1984. Evolution and tertiary structure of proteins. *Ann. Rev. Biophys. Bioeng*，13：453-492

Chothia C，1976. The nature of the accessible and buried surfaces in proteins. *J Mol, Biol.*，**105**：1-14.

Chothia C，Lesk A M，1986. The relation between the divergence of sequence and in proteins *EMBO J.* **5**：823-826.

Creighton T E，1983 Proteins，structures & molecular properties. New York. W. H. Freeman and Company. 118

Efstratiadis A，1980. The structure and evolution of the human β-globin gene family. *Cell*，**21**：653-668.

Hubbard T J P，Blundell T L，1987 Comparison of solvent-inaccessible coress of homologous proteins：Definitions useful for protein modelling *Protein Eng.*，**1**：159-171

Johnson M S，Stucliffe M J，Blundell T L，1990. Molecular anatomy: Phyletic relationships derived from three-dimensional structures of proteins. *J Mol. Evol* **30**：43-59.

Kauzmann W，1959. Some factors in the interpretation of protein denaturation. *Advan. Protein Chem.*，**14**：1-63

Lim W A，Sauer R T，1989. Alternative packing arrangements in the hydrophobic core of λ repressor. Nature，**339**：31-36.

Lee B，Richards F M，1971.The Interpretation of protein structures: estimation of static accessibility. *J Mol. Biol.*，**55**：397--400.

Overington J P，1990 Tertiary structural constraints on protein evolutionary diversity：Templates，key residues and structure prediction. *Proc Roy Soc. London* ser. B **241**：146-152.

Overington J P，1992 Environment-specific amino acid substitution tables：Tertiary templates and prediction of protein folds *Protein Sci*，**1**：216-226

Richards F M，1977. Area，volumes，packing，and protein structure *Ann. Rev. Biophys Bioeng*，**6**：151-176.

Topham C M，1993. Fragment ranking in modelling of protein structure：conformationally-constrained environmental amino acid substitution tables. *J. Mol. Biol*，**229**：194-220.

Wako H，Blundell T L，1994 Use of amino acid environment-dependent substitution tables and conformational protein sites in structure prediction from aligned sequences of homologous proteins 1. Solvent accessibility classes. *J Mol. Biol* **238**：682-692

Wells D E，1986 Compilation analysis of histones and histone genes. *Nucleic Acids Res.*，14：119-149.

AN EXPLORATION OF HISTONE MOLECULAR EVOLUTION FROM THE STATIC ACCESSIBILITY OF AMINO ACID RESIDUES

Huang Jingfei Liu Ciquan

(Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, the Chinese Academy of Sciences, Kunming 650223)

Abstract

This paper investigated and discussed the evolutionary relations of four histone mole-

cules, H_{2A}, H_{2B}, H_3 and H_4 with the static accessibilities of protein amino acid residues as an index. The phylogenetic trees of H_{2A}, H_{2B} and H_3 have been obtained. The results indicated that protein three-dimensional structure is more reliable as studing protein molecular evolution; in addition, H_4 is the most conservative protein in the four histone molecules composing nucleosome. The authors speculate upon that histone H_4, as compared with other histones, perhaps has the most important significance both in keeping chromosomal structure and function and in the origin of chromosome.

Key words Histone, The static accessibility, Molecular evolution, Three-dimensional structure

简 讯

用电生理学方法探索学习记忆的细胞分子机理取得突破性进展

1994年9月，中国科学院昆明动物研究所灵长类生物学研究室学者徐林赴英国伦敦大学药理系进修，1995年6月转入爱尔兰都柏林大学药理和治疗系攻读电生理学博士学位。在此期间，他在学习记忆的细胞分子机理研究方面取得了新突破，研究结果已在《Nature》（Vol. 387／29 May 1997）正式发表。现将其于1997年6月回国讲学交流内容综合简介如下。

1 突触可塑性及其意义

"学习记忆是大脑对信息的获得、储存和提取的过程"(Martinez, 1996)。而突触正是神经细胞之间信息传递的通道。因此，研究突触对神经信息传递的可塑性(long-term potentiation, LTP; long-term depression, LTD)的细胞分子基础，成为深入了解学习记忆机理的关键，同时也是深入了解各种脑疾病（老年性痴呆、精神分裂症、癫痫等）的基础。

2 突触可塑性的研究进展

1973年，Bliss教授首先发现了大脑海马结构中存在着活动依赖性的突触信息传递的长时程增加(LTP)现象。此发现不仅精确地反映了Hebb（1949）提出的学习记忆留下的神经痕迹的假说，而且大大促进了突触信息传递可塑性对学习记忆机理的理解。自1976年诺贝尔奖得主Sakmann发明了膜片钳电生理技术后，神经突触信息传递可塑性研究得以深入到受体离子通道、基因表达、蛋白质合成等细胞分子水平。

随着LTD在海马结构中的发现（Lynch, 1977），大量的报道指出，在各个脑区存在着相同的突触信息传递的可塑性（Linden, 1995）。有的LTP依赖于NMDA受体(Collingridge, 1983)，有的依赖于电压依赖性的 Ca^{2+} 通道（Zalutsky, 1990），也有的依赖于mGlu受体（Nicoll, 1993）。现已知道LTD、LTP突触信息传递的可塑性都依赖于细胞内游离 Ca^{2+} 的增加（Artola, 1993）。

细胞内 Ca^{2+} 的增加引起了一系列分子信息过程的改变。Tonegawa（1996年诺贝尔奖得